

An FPTAS for Minimizing Indefinite Quadratic Forms over Integers in Polyhedra

Robert Hildebrand^{*1} Robert Weismantel^{†1} Kevin Zemmer^{‡1}

¹Department of Mathematics, IFOR, ETH Zürich

October 10, 2015

Abstract

We present a generic approach that allows us to develop a fully polynomial-time approximation scheme (FPTAS) for minimizing nonlinear functions over the integer points in a rational polyhedron in fixed dimension. The approach combines the subdivision strategy of Papadimitriou and Yannakakis [22] with ideas similar to those commonly used to derive real algebraic certificates of positivity for polynomials. Our general approach is widely applicable. We apply it, for instance, to the Motzkin polynomial and to indefinite quadratic forms $\mathbf{x}^T Q \mathbf{x}$ in a fixed number of variables, where Q has at most one positive, or at most one negative eigenvalue. In dimension three, this leads to an FPTAS for general Q .

1 Introduction

Consider the problem

$$(1.1) \quad \min\{f(\mathbf{x}) : \mathbf{x} \in P \cap \mathbb{Z}^n\}$$

for $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $P = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\}$, $A \in \mathbb{Z}^{m \times n}$, and $\mathbf{b} \in \mathbb{Z}^m$. We use the words *size* and *binary encoding length* synonymously. The size of P is the sum of the sizes of A and \mathbf{b} . We say that Problem (1.1) can be solved in polynomial time if in time bounded by a polynomial in the size of its input, we can either determine that the problem is infeasible, find a feasible minimizer, or show that the problem is unbounded by exhibiting a feasible point $\bar{\mathbf{x}}$ and an integer ray $\bar{\mathbf{r}} \in \text{rec}(P) := \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq 0\}$ such that $f(\bar{\mathbf{x}} + \lambda \bar{\mathbf{r}}) \rightarrow -\infty$ as $\lambda \rightarrow \infty$.

The main focus of this paper is Problem (1.1) with $f(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$, where $Q \in \mathbb{Z}^{n \times n}$ is a symmetric matrix. Note that if Q is not symmetric, then we can replace it

by $Q' = \frac{1}{2}Q + \frac{1}{2}Q^T$, which is symmetric and satisfies $\mathbf{x}^T Q \mathbf{x} = \mathbf{x}^T Q' \mathbf{x}$. The input size of Problem (1.1) with $f(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$ is the sum of the sizes of P and Q . When Q is positive semi-definite, $f(\mathbf{x})$ is convex, whereas it is concave when Q is negative semi-definite. In the former case, Problem (1.1) with fixed n and bounded P can be solved in polynomial time by [18], whereas in the latter case by [5] (cf. Theorem 1.2 below).

The computational complexity of mixed-integer polynomial optimization in fixed dimension was surveyed in [8, 7], and they develop an FPTAS for maximizing non-negative polynomials over integer and mixed-integer points in a polytope, respectively. They also use a weaker notion of approximation algorithm, that has also been used in [27], called *fully polynomial-time weak-approximation scheme*. It allows the accuracy to depend on the maximum and minimum values of f on the feasible set. That is, let f_{\max} and f_{\min} be the maximum and minimum values of f on the feasible region. Then a weak ϵ -approximation \mathbf{x}_ϵ of the minimization problem satisfies

$$|f(\mathbf{x}_\epsilon) - f_{\min}| \leq \epsilon(f_{\max} - f_{\min}).$$

When $n \leq 2$ and f is a polynomial of degree d , it has been shown in [11] and [10] that Problem (1.1) can be solved in polynomial time for $d \leq 3$ and in [8] that it is NP-hard for $d = 4$. To the best of our knowledge, the complexity of Problem (1.1) for fixed $n \geq 3$ and when f is a polynomial of degree $d = 2, 3$ is still unknown.

We present an FPTAS for Problem (1.1) with $f(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$ and fixed n in two cases: For Q with at most one negative eigenvalue, and for Q with at most one positive eigenvalue. The numbers of positive, negative, and zero eigenvalues of Q from a triple known as the *inertia* of Q and can be computed in polynomial time (see [3] or [16]).

By [9], for fixed n , it is possible to decide in polynomial time if Problem (1.1) with $f(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$ is

^{*}robert.hildebrand@ifor.math.ethz.ch

[†]robert.weismantel@ifor.math.ethz.ch

[‡]kevin.zemmer@ifor.math.ethz.ch

bounded or not. If it is unbounded, [9] gives a certificate of unboundedness, whereas if it is bounded, it returns an upper bound τ on the size of an optimal solution that is polynomial in the input size. Thus we can replace P by $P \cap \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_\infty \leq 2^\tau\}$. Henceforth we assume that P is bounded.

We use a notion of approximation that is common in combinatorial optimization and is akin to the maximization version used in [8, 7] for maximizing non-negative polynomials over polyhedra, except here, we extend the notion to allow for the objective to take negative values.

DEFINITION 1.1. *Let \mathbf{x}_{\min} be an optimal solution to Problem (1.1) and let $\epsilon > 0$. We say that \mathbf{x}_ϵ is an ϵ -approximate solution to Problem (1.1) if \mathbf{x}_ϵ is feasible and one of the following hold:*

1. $f(\mathbf{x}_{\min}) > 0$ and $f(\mathbf{x}_\epsilon) \leq (1 + \epsilon)f(\mathbf{x}_{\min})$,
2. $f(\mathbf{x}_{\min}) < 0$ and $f(\mathbf{x}_\epsilon) \leq \frac{1}{1+\epsilon}f(\mathbf{x}_{\min})$,
3. $f(\mathbf{x}_{\min}) = 0$ and $f(\mathbf{x}_\epsilon) = f(\mathbf{x}_{\min})$.

We say an algorithm for Problem (1.1) is a fully polynomial-time approximation scheme if for any $\epsilon > 0$, in polynomial time in $\frac{1}{\epsilon}$ and the size of the input, the algorithm correctly determines whether the problem is feasible, and if it is, outputs an ϵ -approximate solution \mathbf{x}_ϵ .

In order to develop an FPTAS for classes of non-linear functions to be minimized over integer points in polyhedra, we propose a framework that combines the techniques of Papadimitriou and Yannakakis [22] with ideas similar to those commonly used to derive certificates of positivity for polynomials over semialgebraic sets. Generally speaking, in the latter context one is given a finite number of “basic polynomials” f_1, \dots, f_m which are known to be positive over the integers in a polyhedron P . A sufficient condition to prove that another polynomial f is positive over $P \cap \mathbb{Z}^n$ is to find a decomposition of f as a sum of products of a *sum of squares* (SOS) polynomial and a basic function f_i . A polynomial $p(\mathbf{x})$ is SOS if there exist polynomials $q_1(\mathbf{x}), \dots, q_m(\mathbf{x})$ such that $p(\mathbf{x}) = \sum_{i=1}^m q_i^2(\mathbf{x})$.

We would like to use a similar approach to arrive at an FPTAS. Again we work with classes of “basic functions”. Then, for a given f , we try to detect a decomposition of f as a finite sum of products of a so-called “sliceable function” and a basic function f_i . Roughly speaking, sliceable functions — thanks to the result of [22] — can be approximated by subdividing the given polyhedron.

For instance, the set of all convex functions presented by a first order oracle that are nonnegative over

$P \cap \mathbb{Z}^n$ could serve as a class of basic functions, because we can solve Problem (1.1) for any member in the class in polynomial time when n is fixed. The nonnegativity assumption implies sign-compatibility, which is a necessary property of the set of basic functions that will be introduced in Section 2. Another example is the set of all concave functions presented by an evaluation oracle that are nonnegative over $P \cap \mathbb{Z}^n$. The same property holds true in this case. These two examples demonstrate that we consider not only polynomials f_i , but also more general classes of basic functions. In fact, this is key to tackle the quadratic optimization problem. For example, we can decompose the polynomial $x^2 + y^2 - z^2$ as the product of two non-polynomial functions: a basic function $\sqrt{x^2 + y^2} - |z|$ and a sliceable function $\sqrt{x^2 + y^2} + |z|$. Our technique also applies, for instance, to the Motzkin polynomial, but to functions f that are not polynomials as well (see Theorem 2.1).

We postpone the discussion of what we mean by basic and sliceable functions to Section 2. As a consequence of our technique we easily derive the following result that in the optimization community was an open question for quite several years.

THEOREM 1.1. *Let $Q \in \mathbb{Z}^{n \times n}$ be a symmetric matrix and let n be fixed. Then there is an FPTAS for Problem (1.1) with $f(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$ in the following cases:*

- (i) Q has at most one negative eigenvalue;
- (ii) Q has at most one positive eigenvalue.

We present a proof for Theorem 1.1 in Section 3. Note that Theorem 1.1 (ii) is equivalent to maximizing $\mathbf{x}^T Q \mathbf{x}$ with at most one negative eigenvalue. Together with known results on quasi-convex and quasi-concave optimization, Theorem 1.1 implies that for $n = 3$ there is an FPTAS for general Q .

COROLLARY 1.1. *Let $Q \in \mathbb{Z}^{n \times n}$ be a symmetric matrix and $n = 3$. Then there is an FPTAS for Problem (1.1) with $f(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$.*

In order to prove this corollary, we need the following known results and notation, which are also used for our other results. Given a convex set C , a function f is *quasi-convex* on C if the set $\{\mathbf{x} \in C : f(\mathbf{x}) \leq \alpha\}$ is convex for all $\alpha \in \mathbb{R}$. Convex functions are quasi-convex on any convex set in their domain. We say that f is *quasi-concave* if $-f$ is quasi-convex. A *semi-algebraic set* in \mathbb{R}^n is a subset of the form $\bigcup_{i=1}^s \bigcap_{j=1}^{r_i} \{\mathbf{x} \in \mathbb{R}^n \mid f_{i,j}(\mathbf{x}) *_{i,j} 0\}$ where $f_{i,j}: \mathbb{R}^n \rightarrow \mathbb{R}$ is a polynomial in n variables and $*_{i,j}$ is either $<$ or $=$ for $i = 1, \dots, s$ and $j = 1, \dots, r_i$ (cf. [1]). For a polyhedron P , denote by P_I the *integer hull* of P , i.e., the convex hull of $P \cap \mathbb{Z}^n$.

THEOREM 1.2. ([18], [5]) *For $A \in \mathbb{Z}^{m \times n}$ and $\mathbf{b} \in \mathbb{Z}^m$, let $P = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\}$ be a bounded polytope, and let n be fixed.*

(i) *If C is a convex, semi-algebraic set given by polynomial inequalities of degree at most d and with integral coefficients of size at most l , and f is a polynomial which is quasi-convex on P , then $\min\{f(\mathbf{x}) : \mathbf{x} \in P \cap C \cap \mathbb{Z}^n\}$ can be solved in polynomial time in d, l and the size of P and the size of the coefficients of all involved polynomials.*

(ii) *The vertices of P_I can be enumerated in polynomial time in the size of P .*

(iii) *If a comparison oracle for f is available, i.e., an oracle that given $\mathbf{x}, \mathbf{y} \in \mathbb{Z}^n$ decides whether $f(\mathbf{x}) < f(\mathbf{y})$ holds or not, and f is quasi-concave on P_I , then Problem (1.1) can be solved in polynomial time in the size of P .*

Proof of Corollary 1.1. The inertia of Q can be computed in polynomial time (see [3] or [16]). If Q has no negative eigenvalues, then f is convex because it is the sum of convex functions, so we can apply Theorem 1.2 (i) with $C = P$. If Q has no positive eigenvalues, then f is concave, so we can apply Theorem 1.2 (iii). If Q has one negative eigenvalue, we can apply Theorem 1.1 (i), whereas if there are two, we can apply Theorem 1.1 (ii). \square

The remainder of this paper is outlined as follows: In Section 2 we explain basic and sliceable functions and present the main tool of this paper. In Section 3 we give some important properties of quadratics and prove Theorem 1.1.

2 Basic and Sliceable Functions

Let \mathcal{P} be a set of polytopes that is closed under taking subsets, i.e., such that for all $\bar{P} \subseteq P$, $P \in \mathcal{P}$ implies $\bar{P} \in \mathcal{P}$. A set of *basic functions* $\Omega_{\mathcal{P}}$ is a class of functions with a common encoding structure for all elements and the following properties:

- (i) $\Omega_{\mathcal{P}}$ is closed under addition and multiplication by $\lambda \geq 0$,
- (ii) each $f \in \Omega_{\mathcal{P}}$ admits a FPTAS for Problem (1.1) over each polytope $P \in \mathcal{P}$,
- (iii) all functions in $\Omega_{\mathcal{P}}$ are sign compatible, i.e., $\text{sign}(f(\mathbf{x})) = \text{sign}(g(\mathbf{x}))$ for all $f, g \in \Omega_{\mathcal{P}}$, $\mathbf{x} \in P$, $P \in \mathcal{P}$.

Examples of classes of functions $\Omega_{\mathcal{P}}$ that satisfy the properties above are the set of sign-compatible

convex function over \mathcal{P} presented by a first order oracle (cf. [14, 21]), sign-compatible concave functions over \mathcal{P} presented by an evaluation oracle (cf. [5]), and the set of polynomials that are negative on some polytope over \mathcal{P} (cf. [8, 7]).

We next introduce a class of functions that can be combined with basic functions and preserve the property that they admit a FPTAS for Problem (1.1). For this, we need the following notation.

For $\mathbf{k} \in \mathbb{Z}^l$, $\mathbf{c}^0 \in \mathbb{Z}^l$, $\mathbf{c}^1, \dots, \mathbf{c}^l \in \mathbb{Z}^n$, and $\epsilon > 0$, define

$$B_{\mathbf{k}}(\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^l, \epsilon) := \left\{ \mathbf{x} \in \mathbb{R}^n : \begin{cases} (1 + \epsilon)^{k_j - 1} \leq L_j(\mathbf{x}) \leq (1 + \epsilon)^{k_j} & \text{if } k_j \geq 1 \\ L_j(\mathbf{x}) = 0 & \text{if } k_j = 0 \\ -(1 + \epsilon)^{-k_j - 1} \leq L_j(\mathbf{x}) \leq -(1 + \epsilon)^{-k_j} & \text{if } k_j \leq -1 \end{cases} \right. \\ \left. \text{for } j = 1, \dots, l \right\}.$$

where $L_j(\mathbf{x}) := \mathbf{c}^j \cdot \mathbf{x} + c_j^0$. Let $B(\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^l, \epsilon) := \bigcup_{\mathbf{k} \in \mathbb{Z}^l} B_{\mathbf{k}}(\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^l, \epsilon)$. Note that for all $\mathbf{c}^0 \in \mathbb{Z}^l$, $\mathbf{c}^1, \dots, \mathbf{c}^l \in \mathbb{Z}^n$, and $\epsilon > 0$ we have that $\mathbb{Z}^n \subseteq B(\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^l, \epsilon)$. However, $\mathbb{R}^n \not\subseteq B(\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^l, \epsilon)$, because $B_{\mathbf{k}}(\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^l, \epsilon)$ is not full-dimensional if any entry of \mathbf{k} is equal to zero.

DEFINITION 2.1. Let $C \subseteq \mathbb{R}^n$. A function $s : C \rightarrow \mathbb{R}_{\geq 0}$ is called *sliceable over C* if there exist $\mathbf{c}^0 \in \mathbb{Z}^l$, $\mathbf{c}^1, \dots, \mathbf{c}^l \in \mathbb{Z}^n$ and $\zeta \in \mathbb{Z}_{\geq 1}$ such that for all $\epsilon > 0$ and for all $\mathbf{k} \in \mathbb{Z}^l$, we have that

$$(2.2) \quad s(\mathbf{x}) \leq (1 + \epsilon)s(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in C \cap B_{\mathbf{k}}(\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^l, \epsilon/\zeta) \cap \mathbb{Z}^n$.

We define the *size of a sliceable function s* as the sum of the sizes of $\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^l$ and the unary encoding size of ζ .

It follows from Definition 2.1 that if a function s is sliceable over C with parameters $\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^l, \zeta$, and $s(\bar{\mathbf{x}}) = 0$ for some $\bar{\mathbf{x}} \in C \cap B_{\mathbf{k}}(\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^l, \epsilon/\zeta) \cap \mathbb{Z}^n$, then $s(\mathbf{x}) = 0$ for all $\mathbf{x} \in C \cap B_{\mathbf{k}}(\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^l, \epsilon/\zeta) \cap \mathbb{Z}^n$. Regardless of this propagation property, there exist large classes of sliceable functions. We present several examples of non-trivial sliceable functions below. To avoid the propagation of the zeros, we give choices of the parameters $\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^l, \zeta$ such that the zeros of the function lay on intersections of hyperplanes defined by $L_j(\mathbf{x}) = 0$ for $j \in J$ for some $J \subseteq 1, \dots, l$.

The set of sliceable functions is not closed under additive inverses, but is in fact a proper semifield as the following lemma shows.

LEMMA 2.1. Let $s, r: C \rightarrow \mathbb{R}_{\geq 0}$ be sliceable functions and let $\lambda \in \mathbb{R}_{\geq 0}$. Then λs , $s + r$ and sr are sliceable with size polynomial in the size of s and r . Moreover, if $s > 0$, then $\frac{1}{s}$ is sliceable with size polynomial in the size of s .

Proof. Let s be sliceable with parameters $\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^{l_s}, \zeta_s$ and r with parameters $\mathbf{d}^0, \mathbf{d}^1, \dots, \mathbf{d}^{l_r}, \zeta_r$. Set $\zeta_{s+r} = \max\{\zeta_s, \zeta_r\}$ and $\zeta_{sr} = 4 \max\{\zeta_s, \zeta_r\}$. We show that $s + r$ and sr are sliceable with the same parameters $[\mathbf{c}^0; \mathbf{d}^0], \mathbf{c}^1, \dots, \mathbf{c}^{l_s}, \mathbf{d}^1, \dots, \mathbf{d}^{l_r}$ and with ζ_{s+r}, ζ_{sr} , respectively. Then for $\zeta = \zeta_{s+r}, \zeta_{sr}$ and $0 < \epsilon < 1$,

$$\begin{aligned} & B_{[\mathbf{k}^1; \mathbf{k}^2]}([\mathbf{c}^0; \mathbf{d}^0], \mathbf{c}^1, \dots, \mathbf{c}^{l_s}, \mathbf{d}^1, \dots, \mathbf{d}^{l_r}, \epsilon/\zeta) \\ & \subseteq B_{\mathbf{k}^1}(\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^{l_s}, \epsilon/\zeta) \cap B_{\mathbf{k}^2}(\mathbf{d}^0, \mathbf{d}^1, \dots, \mathbf{d}^{l_r}, \epsilon/\zeta). \end{aligned}$$

Let $\mathbf{x}, \mathbf{y} \in B_{[\mathbf{k}^1; \mathbf{k}^2]}([\mathbf{c}^0; \mathbf{d}^0], \mathbf{c}^1, \dots, \mathbf{c}^{l_s}, \mathbf{d}^1, \dots, \mathbf{d}^{l_r}, \epsilon/\zeta) \cap \mathbb{Z}^n$. By the containment, it follows that

$$s(\mathbf{x}) \leq (1 + \epsilon \frac{\zeta_s}{\zeta}) s(\mathbf{y}) \leq (1 + \epsilon) s(\mathbf{y})$$

and

$$r(\mathbf{x}) \leq (1 + \epsilon \frac{\zeta_r}{\zeta}) r(\mathbf{y}) \leq (1 + \epsilon) r(\mathbf{y}).$$

For $\zeta = \zeta_{s+r}$, by combining these, we see that $s(\mathbf{x}) + r(\mathbf{x}) \leq (1 + \epsilon)(s(\mathbf{y}) + r(\mathbf{y}))$. Hence, $s + r$ is sliceable. Instead, for $\zeta = \zeta_{sr}$, by combining these, we see that sr is also sliceable since

$$\begin{aligned} s(\mathbf{x})r(\mathbf{x}) & \leq (1 + \epsilon \frac{\zeta_s}{\zeta_{sr}}) s(\mathbf{y}) (1 + \epsilon \frac{\zeta_r}{\zeta_{sr}}) r(\mathbf{y}) \\ & \leq (1 + \frac{\epsilon}{4})^2 s(\mathbf{y}) r(\mathbf{y}) \\ & \leq (1 + \epsilon) s(\mathbf{y}) r(\mathbf{y}). \end{aligned}$$

The function λs is trivially sliceable the same parameters as s . Also, if $s > 0$, then it follows $\frac{1}{s}$ is sliceable with the same parameters as s . Indeed, for $\mathbf{x}, \mathbf{y} \in B_{\mathbf{k}}(\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^{l_s}, \epsilon/\zeta_s) \cap \mathbb{Z}^n$ it holds that $s(\mathbf{y}) \leq (1 + \epsilon) s(\mathbf{x})$. By dividing, we have $\frac{1}{s(\mathbf{x})} \leq (1 + \epsilon) \frac{1}{s(\mathbf{y})}$. \square

Any function that is the sum of even powers of linear forms is sliceable over \mathbb{R}^n . Rational powers of sliceable functions are also sliceable. For instance, $\sqrt{s(\mathbf{x})}$ is sliceable for any sliceable function s . Moreover, any polynomial that is positive on a polyhedron P is sliceable on P . Handelman [15] showed that any polynomial f that is positive over a polyhedron has a representation as $f(\mathbf{x}) = \sum_{|\alpha| \leq L, \alpha \in \mathbb{Z}_+^n} \beta_\alpha \prod_{i=1}^n (\mathbf{a}^i \cdot \mathbf{x} - b_i)^{\alpha_i}$ for some $L \in \mathbb{Z}_+$ and $\beta_\alpha \geq 0$. It follows that f is sliceable over P . Unfortunately, the parameter ζ grows exponentially with L . The best bounds on L are given by Powers and Reznick [24], which are in terms of the coefficients of f and the smallest value that f attains on P . The smallest value is bounded by a polynomial in the input

provided the degree and the dimension are fixed [17], but this still implies that the best known bound for L is not of polynomial size in the input. Therefore there is no known bound on the size of ζ based on the input of the polynomial f that would admit an FPTAS for Problem (1.1) over P with this decomposition.

Example. The polynomial $M(x, y) = x^4 y^2 + x^2 y^4 - 3x^2 y^2 + 1$ is known as the Motzkin polynomial. While it is not SOS, it is known to be positive because $(x^2 + y^2) \cdot M(x, y)$ has an explicit SOS decomposition (see [23, p. 84]). In particular, for $x, y \neq 0$, we have

$$\begin{aligned} M(x, y) &= \frac{1}{(x^2 + y^2)} \left(y^2(1 - x)^2(1 + x)^2 \right. \\ & \quad \left. + x^2(1 - y)^2(1 + y)^2 + x^2 y^2 (x^2 + y^2 - 2)^2 \right). \end{aligned}$$

Thus, for $x, y \neq 0$, $M(x, y)$ is a composition of sliceable functions and a convex function. That is, the parameters $\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^l$ are defined by $c_1^j x + c_2^j y + c_j^0 = F_j(x, y)$ for $F_1(x, y), \dots, F_6(x, y)$ equal to $x, y, 1 - x, 1 + x, 1 - y, 1 + y$ respectively, and the convex function is $(x^2 + y^2 - 2)^2$. Sign-compatibility is also given because all involved functions are non-negative. As we will see in Theorem 2.1, this implies that Problem (1.1) with $f(x, y) = M(x, y)$ admits an FPTAS over polytopes. \square

The following lemma allows us to find approximate solutions by taking the best approximate solution over a decomposition of the feasible region.

LEMMA 2.2. Let $C \subseteq \mathbb{R}^n$ such that $C \cap \mathbb{Z}^n \neq \emptyset$, and for $j = 1, \dots, m$ let $s_j: C \rightarrow \mathbb{R}_{\geq 0}$ for $j = 1, \dots, m$ and let $g_j: C \rightarrow \mathbb{R}$. Let $0 < \epsilon < 1$ and $\epsilon' = \epsilon/4$. Let $K \subseteq \mathbb{Z}^n$ such that $C \cap \mathbb{Z}^n \subseteq \bigcup_{\mathbf{k} \in K} B_{\mathbf{k}}$, where $B_{\mathbf{k}} \subseteq \mathbb{R}^n$ are polytopes. Also, let $L_j^{\mathbf{k}} \geq 0$ such that $L_j^{\mathbf{k}} \leq s_j(\mathbf{x}) \leq (1 + \epsilon') L_j^{\mathbf{k}}$ for all $\mathbf{x} \in B_{\mathbf{k}} \cap \mathbb{Z}^n$ and $j = 1, \dots, m$.

Then, for

$$\mathbf{x}_\epsilon := \arg \min \left\{ \sum_{j=1}^m g_j(\mathbf{x}_{\epsilon'}^{\mathbf{k}}) s_j(\mathbf{x}_{\epsilon'}^{\mathbf{k}}) : \mathbf{k} \in K \right\}$$

where $\mathbf{x}_{\epsilon'}^{\mathbf{k}}$ is an ϵ' -approximate solution to

$$\min \left\{ \sum_{j=1}^m g_j(\mathbf{x}) L_j^{\mathbf{k}} : \mathbf{x} \in C \cap B_{\mathbf{k}} \cap \mathbb{Z}^n \right\},$$

we have that \mathbf{x}_ϵ is an ϵ -approximate solution to

$$\min \left\{ \sum_{j=1}^m g_j(\mathbf{x}) s_j(\mathbf{x}) : \mathbf{x} \in C \cap \mathbb{Z}^n \right\}.$$

Proof. Let

$$\mathbf{x}_{\min} := \arg \min \left\{ \sum_{j=1}^m g_j(\mathbf{x}) s_j(\mathbf{x}) : \mathbf{x} \in C \cap \mathbb{Z}^n \right\},$$

let $\bar{\mathbf{k}} \in K$ such that $\mathbf{x}_{\min} \in B_{\bar{\mathbf{k}}}$, and define $f(\mathbf{x}) := \sum_{j=1}^m g_j(\mathbf{x}) s_j(\mathbf{x})$. Suppose first that $f(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}) \geq 0$, which implies that $g_j(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}) \geq 0$ for all $j = 1, \dots, m$, because all functions in $\Omega_{\mathcal{P}}$ are sign compatible and $s_j \geq 0$. Then

$$\begin{aligned} f(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}) &= \sum_{j=1}^m g_j(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}) s_j(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}) \\ &\leq (1 + \epsilon') \sum_{j=1}^m g_j(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}) L_j^{\mathbf{k}} \\ &\leq (1 + \epsilon')^2 \sum_{j=1}^m g_j(\mathbf{x}_{\min}) L_j^{\mathbf{k}} \\ &\leq (1 + \epsilon')^2 \sum_{j=1}^m g_j(\mathbf{x}_{\min}) s_j(\mathbf{x}_{\min}) \\ &= (1 + \epsilon')^2 f(\mathbf{x}_{\min}) \leq (1 + \epsilon) f(\mathbf{x}_{\min}). \end{aligned}$$

The second inequality follows from approximate minimality of $\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}$, which at the same time implies that $g_j(\mathbf{x}_{\min}) \geq 0$ for all $j = 1, \dots, m$, as otherwise $f(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}) < 0$. In particular, this implies that $f(\mathbf{x}_{\min}) \geq 0$.

Suppose now that $f(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}) < 0$, which implies that $g_j(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}) \leq 0$ for all $j = 1, \dots, m$ because all functions in $\Omega_{\mathcal{P}}$ are sign compatible and $s_j \geq 0$. Then

$$\begin{aligned} f(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}) &= \sum_{j=1}^m g_j(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}) s_j(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}) \leq \sum_{j=1}^m g_j(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}) L_j^{\mathbf{k}} \\ &\leq \frac{1}{1 + \epsilon'} \sum_{j=1}^m g_j(\mathbf{x}_{\min}) L_j^{\mathbf{k}} \\ &\leq \frac{1}{(1 + \epsilon')^2} \sum_{j=1}^m g_j(\mathbf{x}_{\min}) s_j(\mathbf{x}_{\min}) \\ &= \frac{1}{(1 + \epsilon')^2} f(\mathbf{x}_{\min}) \leq \frac{1}{1 + \epsilon} f(\mathbf{x}_{\min}). \end{aligned}$$

To derive those inequalities, we have again used the assumptions on $L_j^{\mathbf{k}}$ and approximate minimality of $\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}}$. In particular, the last inequality holds because $f(\mathbf{x}_{\min}) \leq 0$. The result follows from $f(\mathbf{x}_{\epsilon}) \leq f(\mathbf{x}_{\epsilon'}^{\bar{\mathbf{k}}})$. \square

Let $\mathcal{S}(\mathcal{P})$ be the set of sliceable functions over all $P \in \mathcal{P}$ with an evaluation oracle. We define

$$\mathcal{C}(\Omega_{\mathcal{P}}) := \left\{ \sum_{j=1}^m g_j s_j \mid g_j \in \Omega_{\mathcal{P}}, s_j \in \mathcal{S}(\mathcal{P}), m \text{ finite} \right\}.$$

We now show that there is an FPTAS for all functions in the class $\mathcal{C}(\Omega_{\mathcal{P}})$.

THEOREM 2.1. *Let \mathcal{P} , $\Omega_{\mathcal{P}}$ and $\mathcal{C}(\Omega_{\mathcal{P}})$ be as above. Suppose that there is an evaluation oracle for $f = \sum_{j=1}^m g_j s_j \in \mathcal{C}(\Omega_{\mathcal{P}})$ where $g_j \in \Omega_{\mathcal{P}}$, $s_j \in \mathcal{S}(\mathcal{P})$, $j = 1, \dots, m$. Then for every $P \in \mathcal{P}$, there exists an FPTAS for Problem (1.1), where the size of the input is the sum of the sizes of g_j , s_j for $j = 1, \dots, m$, and P .*

Proof. Let $f = \sum_{j=1}^m g_j s_j \in \mathcal{C}(\Omega_{\mathcal{P}})$, $P \in \mathcal{P}$, and let $\epsilon > 0$. Feasibility of $P \cap \mathbb{Z}^n$ can be determined in polynomial time by [19]. Thus we assume henceforth that $P \cap \mathbb{Z}^n \neq \emptyset$.

Let the slicing parameters of s_j be $\mathbf{c}_j^0, \mathbf{c}_j^1, \dots, \mathbf{c}_j^{l_j}, \zeta_j$ for $j = 1, \dots, m$ and let $\zeta_{\max} = \max_j \{\zeta_j\}$. Since P is bounded, there exists $R \in \mathbb{Z}_{\geq 1}$ of polynomial size in the size of P and the sizes of s_j , $j = 1, \dots, m$, such that $|\mathbf{c}_j^i \cdot \mathbf{x} + c_j^0| \leq R$ for all $\mathbf{x} \in P$, $i = 1, \dots, l_j$, and $j = 1, \dots, m$ (see, for example, [25]). Set $\epsilon' = \frac{\epsilon}{4}$ and $N := \left\lceil \frac{\log_2(R)}{\log_2(1 + \epsilon'/\zeta_{\max})} \right\rceil \leq \left\lceil \log_2(R) \left(\frac{\zeta_{\max}}{\epsilon'} + 1 \right) \right\rceil$. Define

$$B_{\mathbf{k}} := B_{\mathbf{k}}([\mathbf{c}_1^0, \dots, \mathbf{c}_m^0], \mathbf{c}_1^1, \dots, \mathbf{c}_1^{l_1}, \dots, \mathbf{c}_m^1, \dots, \mathbf{c}_m^{l_m}, \epsilon'/\zeta_{\max}).$$

In particular, $\mathbb{Z}^n \subseteq \bigcup_{\mathbf{k} \in \mathbb{Z}^{l_1 + \dots + l_m}} B_{\mathbf{k}}$. Then for $K = \{-N, \dots, N\}^{\sum_{j=1}^m l_j}$ we have $P \cap \mathbb{Z}^n \subseteq \bigcup_{\mathbf{k} \in K} B_{\mathbf{k}}$. Notice that many choices of $\mathbf{k} \in K$ are redundant since each non-empty set $B_{\mathbf{k}}$ is a cell or facet of a cell of the hyperplane arrangement in \mathbb{R}^n given by the hyperplanes $\mathbf{c}_j^i \cdot \mathbf{x} + c_j^0 = 0$ and $\mathbf{c}_j^i \cdot \mathbf{x} + c_j^0 = \pm(1 + \epsilon'/\zeta_{\max})^k$ for $k \in \{0, \dots, N\}$, $i = 1, \dots, l_j$, and $j = 1, \dots, m$. By [26], we can enumerate the $O((2N + 3)^n (\sum_{j=1}^m l_j)^n)$ cells of the hyperplane arrangement in polynomial time with n fixed. Hence, we can find a subset $K' \subseteq K$ with cardinality of polynomial size such that $P \subseteq \bigcup_{\mathbf{k} \in K'} B_{\mathbf{k}}$.

Let \mathbf{x}_{ϵ} be as in Lemma 2.2. For $\mathbf{k} \in K'$, we determine a point $\mathbf{x}^{\mathbf{k}} \in P \cap B_{\mathbf{k}} \cap \mathbb{Z}^n$ using [19]. If no such point exists, then we remove \mathbf{k} from K' . Set $L_j^{\mathbf{k}} := \frac{1}{1 + \epsilon'} s_j(\mathbf{x}^{\mathbf{k}})$, computed using the evaluation oracle for s_j , for $j = 1, \dots, m$. Note that $L_j^{\mathbf{k}}$ satisfies the assumptions of Lemma 2.2 because s_j is sliceable and $\zeta_j \leq \zeta_{\max}$. Then, by Lemma 2.2, \mathbf{x}_{ϵ} is an ϵ -approximate solution to Problem (1.1).

We show that we can compute \mathbf{x}_{ϵ} in polynomial time. For each $\mathbf{k} \in K'$, an ϵ' -approximate solution $\mathbf{x}_{\epsilon'}^{\mathbf{k}}$ to

$$\min \left\{ \sum_{j=1}^m g_j(\mathbf{x}) L_j^{\mathbf{k}} : \mathbf{x} \in P \cap B_{\mathbf{k}} \cap \mathbb{Z}^n \right\}$$

can be computed in polynomial time because $\Omega_{\mathcal{P}}$ is closed under addition and multiplication by positive

scalars, and therefore $\sum_{j=1}^m g_j(\mathbf{x})L_j^{\mathbf{k}} \in \Omega_{\mathcal{P}}$. Finally, $\arg \min\{f(\mathbf{x}_{\epsilon'}^{\mathbf{k}}) : \mathbf{k} \in K'\}$ can be computed in polynomial time using the evaluation oracle of f because $|K'|$ is polynomial in the input size. \square

REMARK 2.1. *The assumptions of Theorem 2.1 can be slightly relaxed if we only consider functions of the form $f(\mathbf{x}) = g_1(\mathbf{x})s_1(\mathbf{x})$, i.e., $\Omega_{\mathcal{P}} = \{\lambda g_1 : \lambda \geq 0\}$ for some $g_1: \mathbb{R}^n \rightarrow \mathbb{R}$ and s_1 sliceable. In fact, in this case no evaluation oracle for s_1 is needed, because an ϵ' -approximate solution to*

$$\min \{g_1(\mathbf{x}) : \mathbf{x} \in P \cap B_{\mathbf{k}} \cap \mathbb{Z}^n\}$$

is also an ϵ' -approximate solution to

$$\min \{g_1(\mathbf{x})L_1^{\mathbf{k}} : \mathbf{x} \in P \cap B_{\mathbf{k}} \cap \mathbb{Z}^n\}$$

for any $L_1^{\mathbf{k}} \geq 0$. Thus, $L_1^{\mathbf{k}}$ does not need to be computed explicitly.

REMARK 2.2. *Theorem 2.1 can be generalized from bounded polyhedra P to more general sets C , provided that for any bounded polyhedron \tilde{P} the feasibility and minimization problems on $C \cap \tilde{P} \cap \mathbb{Z}^n$ and the feasibility problem on $C \cap \mathbb{Z}^n$ can be solved in polynomial time in the size of \tilde{P} and the bit size required to represent C . The class of convex semi-algebraic sets in fixed dimension is an example of such a class, because for fixed dimension the stated feasibility problem can be solved in polynomial time by [18].*

REMARK 2.3. *The size of a sliceable function s is defined as the sum of the sizes of its slicing parameters, where unary encoding is assumed for the parameter ζ . With this definition, a polynomial bound on the cardinality of K' in terms of the sizes of the sliceable functions is derived in the proof of Theorem 2.1. This bound is not polynomial if binary encoding is assumed for the ζ parameters. However, in our main application of Theorem 2.1, namely the proof of Theorem 1.1, those parameters can be chosen to be constant.*

3 FPTAS for Quadratic Forms

For positive definite matrices, it is well known that a Cholesky decomposition $Q = LDL^T$ can be easily computed, where L is a lower-triangular matrix and D is a diagonal matrix. Such a decomposition does not always exist for indefinite symmetric matrices [6]. However, there is a similar decomposition that serves our purpose, as described in the following remark.

REMARK 3.1. ([6, 16]) *For any symmetric matrix $Q \in \mathbb{Z}^{n \times n}$, we can find an invertible matrix $S \in \mathbb{Q}^{n \times n}$ and a diagonal matrix $D \in \mathbb{Q}^{n \times n}$ such that $Q = SDS^T$ in polynomial time in the size of Q for fixed n .*

We thus have

$$\mathbf{x}^T Q \mathbf{x} = \sum_{i=1}^n d_i (S_i \cdot \mathbf{x})^2$$

where $D = \text{diag}(\mathbf{d})$ is the diagonal matrix with $D_{ii} = d_i$, and S_i denotes the i -th row of S .

REMARK 3.2. *Since an FPTAS for $c^3 f(\mathbf{x})$, $c \in \mathbb{Q}_{>0}$, is also an FPTAS for $f(\mathbf{x})$ as long as c is of polynomial size in the input size, S and D can be scaled to be in $\mathbb{Z}^{n \times n}$. In fact, let c be the least common multiple of the denominators of all entries of S and D , which is of polynomial size in the size of Q , because S and D can be computed in polynomial time in the size of Q . Then $c^3 f(\mathbf{x}) = \mathbf{x}^T (cS)(cD)(cS)^T \mathbf{x}$ and $cS, cD \in \mathbb{Z}^{n \times n}$.*

Since S is non-singular, Sylvester's law of inertia (see, for example [13] or [12]) states that the inertia of Q is the same as the inertia of D . Therefore D has exactly as many positive, negative and zero diagonal entries as Q has positive, negative and zero eigenvalues.

The following lemmas will be useful for determining the accuracy necessary for the numerical approximations.

LEMMA 3.1. (SEE, E.G., [10]) *Let $m < n$ be nonnegative integers, $a_m, \dots, a_n \in \mathbb{R}$, $a_m \neq 0$, $a_n \neq 0$, and let $\bar{x} \in \mathbb{R}$ be a nonzero root of the polynomial $f(x) := \sum_{i=m}^n a_i x^i$. Then $|\bar{x}| > \min\{|a_m|/(|a_m| + |a_i|) : i = m+1, \dots, n\}$.*

The proof of the next lemma uses a standard technique for converting an equation with rational powers to one with integer powers.

LEMMA 3.2. *Let $p, q, p', q' \in \mathbb{Z}$ with $|p|, |q|, |p'|, |q'| \leq R$, $p, p' \geq 0$, and let $\delta = (\sqrt{p} + q) - (\sqrt{p'} + q')$. If $\delta \neq 0$, then $|\delta| > \frac{1}{5.48} R^{-4}$.*

Proof. We show that any solution δ is a solution to a polynomial equation in terms of δ . Rewriting the equation for δ by solving for \sqrt{p} and squaring both sides yields $p = \delta^2 + p' + q^2 - 2qq' + q'^2 + \delta(-2q + 2q') + \sqrt{p'}(2\delta - 2q + 2q')$. Solving for $\sqrt{p'}(2\delta - 2q + 2q')$ and squaring both sides yields a degree 4 polynomial in δ with integer coefficients that are polynomial in R . In particular, it can be shown the the absolute values of the coefficients are upper bounded by $48R^4$. Since δ is a solution to this polynomial equation, by Lemma 3.1, it follows that if $\delta \neq 0$, then $|\delta| > \frac{1}{5.48} R^{-4}$. \square

With those technical tools, we can now prove Theorem 1.1.

Proof of Theorem 1.1. Given $Q \in \mathbb{Z}^{n \times n}$, it is possible to compute S and $D = \text{diag}(\mathbf{d})$ such that $Q = SDS^T$ in polynomial time by Remark 3.1. Moreover, S and D can be assumed to be integral by Remark 3.2. Suppose that Q has at most one negative eigenvalue, which implies that $-Q$ has at most one positive eigenvalue. By reordering variables, we may assume that $d_i \geq 0$ for $i = 1, \dots, n-1$. Note that if $d_n \geq 0$, then $f_1(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$ is convex and $f_2(\mathbf{x}) = \mathbf{x}^T (-Q) \mathbf{x}$ is concave, so Problem (1.1) with $f = f_1$ and $f = f_2$ can be solved in polynomial time by Theorem 1.2 (i) and (iii) respectively. Thus assume henceforth that $d_n < 0$.

Let

$$g(\mathbf{x}) := \sqrt{\sum_{i=1}^{n-1} d_i (S_i \cdot \mathbf{x})^2} - \sqrt{-d_n} |S_n \cdot \mathbf{x}|$$

and

$$s(\mathbf{x}) := \sqrt{\sum_{i=1}^{n-1} d_i (S_i \cdot \mathbf{x})^2} + \sqrt{-d_n} |S_n \cdot \mathbf{x}|.$$

Notice that $f_1(\mathbf{x}) = g(\mathbf{x})s(\mathbf{x})$ and $f_2(\mathbf{x}) = (-g(\mathbf{x}))s(\mathbf{x})$. Let $\mathcal{P} := \{\bar{P} \text{ polytope} : \bar{P} \subseteq P\}$ and $\Omega_{\mathcal{P}} := \{\lambda g : \lambda \in \mathbb{R}\}$. We will show that s is sliceable with size bounded by a polynomial in the size of Q , and that $\Omega_{\mathcal{P}}$ is a class of functions that we can minimize over the integers in polyhedra $\bar{P} \in \mathcal{P}$. Equivalently, we show that we can minimize g and $-g$ over $\bar{P} \cap \mathbb{Z}^n$ for $\bar{P} \in \mathcal{P}$. The result then follows from Theorem 1.1 and Remark 2.1.

First, we define the slicing parameters $\mathbf{c}^j = S_j$ for $j = 1, \dots, n$, $\mathbf{c}^0 = 0$, and $\zeta = 1$. Then for any $\mathbf{k} \in \mathbb{Z}^n$, $\epsilon > 0$, and any $\mathbf{x}, \mathbf{y} \in B_{\mathbf{k}}(\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^n, \epsilon)$, we have

$$\begin{aligned} s(\mathbf{x}) &= \sqrt{\sum_{i=1}^{n-1} d_i (S_i \cdot \mathbf{x})^2} + \sqrt{-d_n} |S_n \cdot \mathbf{x}| \\ &\leq \sqrt{\sum_{i=1}^{n-1} d_i ((1+\epsilon) S_i \cdot \mathbf{y})^2} + \sqrt{-d_n} |(1+\epsilon) S_n \cdot \mathbf{y}| \\ &\leq (1+\epsilon) s(\mathbf{y}). \end{aligned}$$

Hence, s is sliceable with size polynomial in the size of Q .

Second, we show that we can solve $\min\{g(\mathbf{x}) : \mathbf{x} \in \bar{P} \cap \mathbb{Z}^n\}$ exactly for any $\bar{P} \in \mathcal{P}$. The function g is convex on the halfspaces $S_n \cdot \mathbf{x} \geq 0$ and $S_n \cdot \mathbf{x} \leq 0$, because it is the sum of a norm with a linear function. Even though g is convex, an exact first order oracle is not available because the function values can be irrational. We show the minimization problem can be solved by reformulating sublevel sets as convex semi-algebraic sets.

We separately consider the problem on the halfspaces $S_n \cdot \mathbf{x} \geq 0$ and $S_n \cdot \mathbf{x} \leq 0$. The analysis for both cases is similar. Without loss of generality, assume that $S_n \cdot \mathbf{x} \geq 0$.

For $\beta \in \mathbb{R}$, the inequality $g(\mathbf{x}) \leq \frac{\beta}{\sqrt{-d_n}}$ is equivalent to

$$\sqrt{|d_n| \sum_{i=1}^{n-1} d_i (S_i \cdot \mathbf{x})^2} \leq \beta + |d_n| S_n \cdot \mathbf{x}.$$

This is a second order cone constraint, which is well known to define a convex set [2]. It can be reformulated by recognizing that the left hand side is nonnegative and adding the condition that the right hand side is nonnegative too. With this, we can square both sides to get the equivalent relation

$$(3.3) \quad |d_n| \sum_{i=1}^{n-1} d_i (S_i \cdot \mathbf{x})^2 - (\beta + |d_n| S_n \cdot \mathbf{x})^2 \leq 0, \\ \beta + |d_n| S_n \cdot \mathbf{x} \geq 0.$$

In this representation, the set $\{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) \leq \frac{\beta}{\sqrt{-d_n}}, S_n \cdot \mathbf{x} \geq 0\}$ is a convex semi-algebraic set. It is thus possible to test integer feasibility of this set for $\beta \in \mathbb{Q}$ in polynomial time using Theorem 1.2 (i). Since \bar{P} is bounded, and S and D are of polynomial size, it follows from [25] that we can compute $R \in \mathbb{Z}$, $R \geq \|\mathbf{d}\|_{\infty}$ of polynomial size such that $\|S\mathbf{x}\|_{\infty} \leq R$ for all $\mathbf{x} \in \bar{P}$. Note that $|\sqrt{-d_n} g(\mathbf{x})| \leq 2nR^3$ for all $\mathbf{x} \in \bar{P}$. We now perform binary search on β in the interval $[-2nR^3, 2nR^3]$ for integer feasibility of (3.3) with precision $\mu := \frac{1}{5.48} n^{-4} R^{-16}$. Since these numbers are of polynomial size, the binary search can be done in polynomial time. We show now that the choice of precision yields an optimal solution by showing that it distinguishes function values of $\sqrt{-d_n} g(\mathbf{x})$ for $\mathbf{x} \in \bar{P} \cap \mathbb{Z}^n$.

Indeed, consider $\mathbf{x}, \mathbf{y} \in \bar{P} \cap \mathbb{Z}^n$ and define $\delta := \sqrt{-d_n}(g(\mathbf{x}) - g(\mathbf{y}))$. Setting $p = |d_n| \sum_{i=1}^{n-1} d_i (S_i \cdot \mathbf{x})^2$, $q = -|d_n| S_n \cdot \mathbf{x}$, $p' = |d_n| \sum_{i=1}^{n-1} d_i (S_i \cdot \mathbf{y})^2$, and $q' = -|d_n| S_n \cdot \mathbf{y}$, we see that $\delta = (\sqrt{p} + q) - (\sqrt{p'} + q')$. Since $p, p' \geq 0$ and $|p|, |p'|, |q|, |q'| \leq nR^4$, by Lemma 3.2, if $\delta \neq 0$, then $|\delta| > \mu$. Thus our choice of the precision for the binary search distinguishes function values of $\sqrt{-d_n} g(\mathbf{x})$ for $\mathbf{x} \in \bar{P}$ and hence we arrive at an exact algorithm that runs in polynomial time. Note also, that this implies that the precision μ/R distinguishes function values of $g(\mathbf{x})$ for $\mathbf{x} \in \bar{P}$.

Lastly, we show that we can solve $\min\{-g(\mathbf{x}) : \mathbf{x} \in \bar{P} \cap \mathbb{Z}^n\}$ for any $\bar{P} \in \mathcal{P}$. Since g is convex, $-g$ is concave and hence is also quasi-concave. By Theorem 1.2 (iii), the minimization problem can be solved exactly provided that we have a comparison oracle for $-g$. Our comparison oracle is realized by approximating $-g$ to a

precision of μ/R , which, as stated above, distinguishes function values of g . This precision can be obtained by approximating the square roots in g using, for instance binary search. See also [20] for approximation of square roots.

Thus, we have shown that the set $\Omega_{\mathcal{P}}$ is indeed a class of functions that we can minimize in polynomial time over polytopes $\bar{P} \in \mathcal{P}$. \square

REMARK 3.3. *In the proof of Theorem 1.1, there is an alternative method that avoids computing solutions to $\min\{-g(\mathbf{x}) : \mathbf{x} \in \bar{P} \cap \mathbb{Z}^n\}$, thus avoiding potential numerical issues when approximating square roots. In fact, let $\bar{\mathbf{x}} \in \arg\min\{f(\mathbf{x}) : \mathbf{x} \text{ is a vertex of } (\bar{P} \cap \mathbb{Z}^n)_I\}$ and $\tilde{\mathbf{x}} \in \arg\min\{-g(\mathbf{x}) : \mathbf{x} \in \bar{P} \cap \mathbb{Z}^n\}$. Since $-g$ is concave, its optimal value over $\bar{P} \cap \mathbb{Z}^n$ is attained at a vertex of $(\bar{P} \cap \mathbb{Z}^n)_I$. Therefore, because $\bar{\mathbf{x}}$ minimizes f over the vertices of $(\bar{P} \cap \mathbb{Z}^n)_I$, we have $f(\bar{\mathbf{x}}) \leq f(\tilde{\mathbf{x}})$.*

Finally, we mention that for certain cases, we can obtain exact solutions to Problem (1.1) for $f(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$ where Q has at most one negative or at most one positive eigenvalue. For this, we use the following lemma.

LEMMA 3.3. *Let $S \in \mathbb{Z}^{n \times n}$ be non-singular and $D = \text{diag}(\mathbf{d})$ with $\mathbf{d} \in \mathbb{Z}_{\geq 0}^{n-1} \times \mathbb{Z}_{< 0}$. Then $f(\mathbf{x}) := \mathbf{x}^T S D S^T \mathbf{x}$ is quasi-convex on the sets $C_1 := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq 0, S_n \cdot \mathbf{x} \geq 0\}$ and $C_2 := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq 0, S_n \cdot \mathbf{x} \leq 0\}$.*

Proof. We show this for the set C_1 ; the proof for the set C_2 is similar. We only need to consider the case where S is the identity matrix, because convexity is preserved under linear transformations.

$C_1 = \{\mathbf{x} \in \mathbb{R}^n : \sqrt{\sum_{i=1}^{n-1} d_i(x_i)^2} \leq \sqrt{-d_n} x_n, x_n \geq 0\}$ is a convex because it is a second-order cone [2]. In particular, its interior, which is given by $\text{int}(C_1) = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) < 0, x_n > 0\}$, is convex.

Note that if $d_i = 0$, then f is quasi-convex on some set C if and only if the restriction of f to $x_i = 0$ is quasi-convex on $C \cap \{\mathbf{x} \in \mathbb{R}^n : x_i = 0\}$. Therefore, by reduction of variables, assume without loss of generality that $d_i \neq 0$ for $i = 1, \dots, n$.

Since $d_i \neq 0$ for $i = 1, \dots, n$, C_1 is a pointed cone and therefore does not contain any lines. Let $\mathbf{x}, \mathbf{y} \in \text{int}(C_1)$ and $\lambda \in (0, 1)$. We will show that $f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \leq \max\{f(\mathbf{x}), f(\mathbf{y})\}$, thus proving quasi-convexity of f on $\text{int}(C_1)$. Let

$$\begin{aligned} g(\lambda) &:= f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \\ &= \lambda^2 f(\mathbf{x}) + (1 - \lambda)^2 f(\mathbf{y}) \\ &\quad + 2\lambda(1 - \lambda) \sum_{i=1}^n d_i (S_i \cdot \mathbf{x})(S_i \cdot \mathbf{y}), \end{aligned}$$

and note that g is a quadratic function in λ and satisfies $g(0) = f(\mathbf{x})$ and $g(1) = f(\mathbf{y})$. If g is convex, then the result follows by convexity. Otherwise, g is strictly concave. Assume without loss of generality that $f(\mathbf{x}) \geq f(\mathbf{y})$. We will show that g is increasing on the interval $[0, 1]$, thus proving the result.

Assume for the sake of contradiction that g is not increasing on the interval $[0, 1]$ and therefore has a local maximum on this interval, say for $\hat{\lambda}$. Since g is quadratic and strictly concave, $\hat{\lambda}$ is also its global maximum. By convexity of $\text{int}(C_1)$, $\hat{\lambda} \mathbf{x} + (1 - \hat{\lambda}) \mathbf{y} \in \text{int}(C_1)$ and thus $g(\hat{\lambda}) < 0$. However, $\text{int}(C_1)$ does not contain any line. Therefore, there is a $\bar{\lambda} \in \mathbb{R}$ such that $\bar{\lambda} \mathbf{x} + (1 - \bar{\lambda}) \mathbf{y} \notin \text{int}(C_1)$, implying $f(\bar{\lambda} \mathbf{x} + (1 - \bar{\lambda}) \mathbf{y}) \geq 0$ because f takes the value 0 on the boundary of C_1 . By maximality of $g(\hat{\lambda})$, this implies that $g(\hat{\lambda}) \geq 0$, a contradiction. Therefore, f is quasi-convex on $\text{int}(C_1)$.

Finally, if a continuous function is continuous on the closure of a convex set C and quasi-convex on its interior, it is quasi-convex on the closure of C (see, for example, Theorem 2.2.12 in [4]). Therefore f is quasi-convex in C_1 . \square

PROPOSITION 3.1. *Let $Q \in \mathbb{Z}^{n \times n}$ be a symmetric matrix and let n be fixed. Then there is a polynomial time algorithm to solve Problem (1.1) with $f(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x}$ in the following cases:*

- (i) Q has exactly one negative eigenvalue and $f(\mathbf{x}_{\min}) \leq 0$;
- (ii) Q has exactly one positive eigenvalue and $f(\mathbf{x}_{\min}) \geq 0$.

Furthermore, in polynomial time, we can detect if one of these cases occurs.

Proof. We suppose that $P \cap \mathbb{Z}^n \neq \emptyset$. As mentioned before, we can determine the inertia of Q in polynomial time.

Suppose Q has exactly one negative eigenvalue. Consider the sets C_1 and C_2 as in Lemma 3.3. Since f is quasi-convex on each of those sets, we can minimize f over $C_1 \cap P \cap \mathbb{Z}^n$ and $C_2 \cap P \cap \mathbb{Z}^n$ in polynomial time by Theorem 1.2 (i). As $C_1 \cup C_2 = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq 0\}$, we have shown how to detect if $f(\mathbf{x}_{\min}) \leq 0$ and if so, determine an optimal solution.

Suppose instead that Q has exactly one positive eigenvalue. We consider the decomposition in Lemma 3.3 for $-Q = S(-D)S^T$, which has exactly one negative eigenvalue. Since we decomposed $-Q$ instead of Q , the sets C_1, C_2 are $C_1 := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \geq 0, S_n \cdot \mathbf{x} \geq 0\}$ and $C_2 := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \geq 0, S_n \cdot \mathbf{x} \leq 0\}$. Now we compute the vertices of the integer hulls $P_I^1 := (P \cap \{\mathbf{x} \in \mathbb{R}^n : S_n \cdot \mathbf{x} \geq 0\})_I$ and

$P_I^2 := (P \cap \{\mathbf{x} \in \mathbb{R}^n : S_n \cdot \mathbf{x} \leq 0\})_I$, which can be done in polynomial time by Theorem 1.2 (ii). A standard convexity argument shows that $f(\mathbf{x}_{\min}) < 0$ if and only if there exists a vertex \mathbf{x}_v of P_I^1 or P_I^2 such that $f(\mathbf{x}_v) < 0$. Otherwise, all vertices of P_I^1 and P_I^2 are contained in $C_1 \cup C_2$. This certifies that $P \cap \mathbb{Z}^n \subseteq C_1 \cup C_2$, and hence $f(\mathbf{x}_{\min}) \geq 0$. If this happens, f is quasi-concave on both C_1 and C_2 . By Theorem 1.2 (iii), we can compute an optimal solution in polynomial time. \square

References

- [1] J. Bochnak, M. Coste, and M.-F. Roy. *Real Algebraic Geometry*. Number 36 in *Ergebnisse der Mathematik und ihrer Grenzgebiete / A Series of Modern Surveys in Mathematics*. Springer Berlin Heidelberg, Jan. 1998.
- [2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.
- [3] J. Bunch. Partial Pivoting Strategies for Symmetric Matrices. *SIAM Journal on Numerical Analysis*, 11(3):521–528, June 1974.
- [4] A. Cambini and L. Martein. *Generalized Convexity and Optimization: Theory and Applications*, volume 616 of *Lecture Notes in Economics and Mathematical Systems*. Springer Berlin Heidelberg, 2009.
- [5] W. J. Cook, M. E. Hartmann, R. Kannan, and C. McDiarmid. On integer points in polyhedra. *Combinatorica*, 12(1):27–37, 1992.
- [6] A. Dax and S. Kaniel. Pivoting techniques for symmetric Gaussian elimination. *Numerische Mathematik*, 28(2):221–241, June 1977.
- [7] J. De Loera, R. Hemmecke, M. Köppe, and R. Weismantel. FPTAS for optimizing polynomials over the mixed-integer points of polytopes in fixed dimension. *Mathematical Programming*, 115(2):273–290, 2008.
- [8] J. A. De Loera, R. Hemmecke, M. Köppe, and R. Weismantel. Integer polynomial optimization in fixed dimension. *Mathematics of Operations Research*, 31:147–153, 2006.
- [9] A. Del Pia, S. S. Dey, and M. Molinaro. Mixed-integer quadratic programming is in NP. Manuscript, 2014.
- [10] A. Del Pia, R. Hildebrand, R. Weismantel, and K. Zimmer. Minimizing cubic and homogeneous polynomials over integers in the plane. *Mathematics of Operations Research*, 2015. accepted for publication.
- [11] A. Del Pia and R. Weismantel. Integer quadratic programming in the plane. In *Proceedings of SODA 2014*, pages 840–846, 2014.
- [12] P. E. Gill, W. Murray, and M. H. Wright. *Numerical Linear Algebra and Optimization*, volume 1. Addison-Wesley Publishing Company, Redwood City, CA, 1991.
- [13] G. H. Golub and C. F. V. Loan. *Matrix Computations*. JHU Press, 3 edition, Oct. 1996.
- [14] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag Berlin Heidelberg, 1988.
- [15] D. Handelman. Representing polynomials by positive linear functions on compact convex polyhedra. *Pacific J. Math.*, 132(1):35–62, 1988.
- [16] R. Hartung. *Computational Problems of Quadratic Forms: Complexity and Cryptographic Perspectives*. PhD thesis, Johann Wolfgang Goethe-Universität, Frankfurt am Main, 2007.
- [17] G. Jeronimo and D. Perrucci. On the minimum of a positive polynomial over the standard simplex. *Journal of Symbolic Computation*, 45(4):434 – 442, 2010.
- [18] L. Khachiyan and L. Porkolab. Integer optimization on convex semialgebraic sets. *Discrete and Computational Geometry*, 23(2):207–224, 2000.
- [19] H. W. Lenstra, Jr. Integer programming with a fixed number of variables. *Mathematics of Operations Research*, 8:538–548, 1983.
- [20] Y. Mansour, B. Schieber, and P. Tiwari. The complexity of approximating the square root. In *30th Annual Symposium on Foundations of Computer Science, 1989*, pages 325–330, Oct. 1989.
- [21] T. Oertel, C. Wagner, and R. Weismantel. Integer convex minimization by mixed integer linear optimization. *Operations Research Letters*, 42(6–7):424 – 428, 2014.
- [22] C. H. Papadimitriou and M. Yannakakis. On the approximability of trade-offs and optimal access of web sources. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science, FOCS '00*, pages 86–92, Washington, DC, USA, 2000. IEEE Computer Society.
- [23] P. Parrilo. Polynomial Optimization, Sums of Squares, and Applications. In *Semidefinite Optimization and Convex Algebraic Geometry*, MOS-SIAM Series on Optimization, pages 47–157. Society for Industrial and Applied Mathematics, Dec. 2012.
- [24] V. Powers and B. Reznick. A new bound for Pólya’s Theorem with applications to polynomials positive on polyhedra. *Journal of Pure and Applied Algebra*, 164(1–2):221–229, Oct. 2001.
- [25] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley and Sons, New York, 1986.
- [26] N. Sleumer. Output-sensitive cell enumeration in hyperplane arrangements. In S. Arnborg and L. Ivansson, editors, *Algorithm Theory – SWAT’98*, volume 1432 of *Lecture Notes in Computer Science*, pages 300–309. Springer Berlin Heidelberg, 1998.
- [27] S. A. Vavasis. Approximation algorithms for indefinite quadratic programming. *Mathematical Programming*, 57(1-3):279–311, May 1992.